# APPLICATIONS OF GROWTH CURVE PREDICTION

by

Jack C. Lee and Seymour Geisser

Technical Report No. 180

August 1972

University of Minnesota
Minneapolis, Minnesota

## 1. Introduction.

In this paper we continue the study of partial or conditional prediction from the growth curve model. Previously we (1972) presented theoretical investigations of this problem and now we apply these results focusing on an empirical or data analytic point of view. The model considered is

$$E(Y_{p\times N}) = X_{p\times m}\tau_{m\times r}A_{r\times N}$$

where $\tau$ is unknown, $X$ and $A$ are known design matrices of ranks $m < p$ and $r < N$ respectively. Further, the columns of $Y$ are independent and p-dimensional multinormal variates having a common unknown covariance matrix $\Sigma$.

For conditional prediction we assume that after observing the sample $Y$, some partial observations on $V$, namely $V^{(1)}$, are also at hand and our interest is in predicting $V^{(2)}$ given $Y$ and $V^{(1)}$ where

$$V_{p\times K} = \begin{array}{c} p_1 \\ p_2 \end{array} \begin{pmatrix} V^{(1)} \\ V^{(2)} \end{pmatrix}, \quad p_1 + p_2 = p.$$

We assume that $V$ is drawn from the growth curve model such that $E(V|\tau, \Sigma) = X\tau F$ where $F_{r\times K}$ is a known design matrix, usually formed by some columns of $A$, and the columns of $V$, given $\tau$ and $\Sigma$, are independent and p-dimensional multivariate normal with common covariance matrix $\Sigma$. Utilizing a Bayesian analysis, we (1972) derived the requisite distributions for predicting $V^{(2)}$ given $V^{(1)}$ and $Y$.

We shall now present an empirical comparison of the relative merits of several predictors for predicting $V^{(2)}$ for various covariance matrix models. In the comparison of predictors we utilize the data given by Potthoff and Roy (1964) and Grizzle and Allen (1969), viz. the 27 vectorial observations in a dental study and the 20 vectorial observations in a ramus height study. In the prediction study the samples are of size $4 \times 26$ and $4 \times 19$ respectively and are different for each $V$ predicted, i.e., the samples are all of the observations except the vector predicted, and we shall predict the last component given the first three and the previous sample, i.e., $p_1 = 3$, $p_2 = 1$. In Section 2 we shall list the predictors to be compared and in Sections 3 and 4 we compare the predictors based on the aforementioned data sets. In Section 5 we make some concluding remarks and suggestions for handling other data sets.

## 2. Conditional Predictors.

In what follows, predictors (A) - (D) were developed by Lee and Geisser (1972) and (E) - (I) are some heuristic predictors that may have some value.

A) Approximate Mean (A.ME.) for arbitrary positive definite (p.d.) $\Sigma$.

When $\Sigma$ is arbitrary p.d., the approximate mean of $V^{(2)}$ given $V^{(1)}$, obtainable from (2.11) of (1972), is

$$\mu_{a2\cdot 1} = \mu_a^{(2)} + \Sigma_{a21}\Sigma_{a11}^{-1}(V^{(1)} - \mu_a^{(1)}) \tag{2.1}$$

where

$$\left\{\begin{array}{l} X = \binom{X^{(1)}}{X^{(2)}} \\ \mu_a^{(i)} = X^{(i)}(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}YA'(AA')^{-1}F, \ i = 1,2 \\ \\ S = Y(I - A'(AA')^{-1}A)Y' \\ \Sigma_a = \begin{pmatrix} \Sigma_{a11} & \Sigma_{a12} \\ \Sigma_{a21} & \Sigma_{a22} \end{pmatrix} \\ \Sigma_a = X(X'\hat{\Sigma}^{-1}X)^{-1}X'M^{-1} + XB\hat{\Sigma}Z(Z'\hat{\Sigma}Z)^{-1}Z'\hat{\Sigma}B'X' + D'Z'\hat{\Sigma}B'X + XB\hat{\Sigma}ZD + D'Z'\hat{\Sigma}ZD \\ M = 1 - F'(HH')^{-1}F \\ H = (A, F) \\ D = (Z'Z)^{-1}Z' \\ B = (X'X)^{-1}X' \end{array}\right. \tag{2.2}$$

and $\hat{\Sigma}$ is some estimate of $\Sigma$. In this study we employ an unbiased estimate of $\Sigma$ for $\hat{\Sigma}$, namely, $\hat{\Sigma} = (N-r)^{-1}S$.

B) Mode for the arbitrary p.d. $\Sigma$.

The mode is the value $V_m^{(2)}$ which maximizes

$$
\begin{aligned}
(2.3)\qquad f(V^{(2)}|V^{(1)}) \propto\ & [K_1 + (V^{(2)} - \mu_{2\cdot1})'\, Q_{22}^{(1)}(V^{(2)} - \mu_{2\cdot1})]^{-(N+1-m)/2} \\
& \cdot[K_2 + (V^{(2)} - \tilde{\mu}_{2\cdot1})'Q_{22}^{(2)}(V^{(2)} - \tilde{\mu}_{2\cdot1}) \\
& \quad + (V^{(2)} - \dot{\mu}_{2\cdot1})'Q_{22}^{(3)}(V^{(2)} - \dot{\mu}_{2\cdot1})]^{-(N+1-r)/2} \\
& \cdot[K_3 + (V^{(2)} - \tilde{\mu}_{2\cdot1})'Q_{22}^{(2)}(V^{(2)} - \tilde{\mu}_{2\cdot1})]^{(N+1-r-m)/2}
\end{aligned}
$$

where

$$
(2.4)\qquad \begin{cases}
K_1 = 1 + (V^{(1)} - X^{(1)}\hat{\tau}F)'Q_{11\cdot2}^{(1)}(V^{(1)} - X^{(1)}\hat{\tau}F) \\
\mu_{2\cdot1} = X^{(2)}\hat{\tau}F - Q_{22}^{(1)-}Q_{21}^{(1)}(V^{(1)} - X^{(1)}\hat{\tau}F) \\
K_2 = M^{-1} + (V^{(1)} - \hat{V}^{(1)})'Q_{11\cdot2}^{(2)}(V^{(1)} - \hat{V}^{(1)}) + (V^{(1)} - X^{(1)}\hat{\tau}F)'Q_{11\cdot2}^{(3)}(V^{(1)} - X^{(1)}\hat{\tau}F) \\
\hat{V} = \begin{pmatrix} \hat{V}^{(1)} \\ \hat{V}^{(2)} \end{pmatrix} \\
X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \\
\tilde{\mu}_{2\cdot1} = \hat{V}^{(2)} - Q_{22}^{(2)-}Q_{21}^{(2)}(V^{(1)} - \hat{V}^{(1)}) \\
Z = \begin{pmatrix} Z^{(1)} \\ Z^{(2)} \end{pmatrix} \\
\dot{\mu}_{2\cdot1} = X^{(2)}\hat{\tau}F - Q_{22}^{(3)-}Q_{21}^{(3)}(V^{(1)} - X^{(1)}\hat{\tau}F) \\
K_3 = M^{-1} + (V^{(1)} - \hat{V}^{(1)})'Q_{11\cdot2}^{(2)}(V^{(1)} - \hat{V}^{(1)}) \\
Q^{(i)} = \begin{pmatrix} Q_{11}^{(i)} & Q_{12}^{(i)} \\ Q_{21}^{(i)} & Q_{22}^{(i)} \end{pmatrix} \\
Q_{11\cdot2}^{(i)} = Q_{11}^{(i)} - Q_{12}^{(i)}Q_{22}^{(i)-}Q_{21}^{(i)} ,
\end{cases}
$$

and $G^-$ is a g-inverse of G satisfying $GG^-G = G$ with the convention that $G^- = 0$ if $G = 0$. The predictive density (2.3) was derived previously by the authors (1972). It was also demonstrated there that the mode could be obtained by a 2-dimensional search procedure. However when $p_2 = 1$, it is clear that we may plot (2.3) and obtain the mode directly.

C) Approximate Mean (A.ME.) for the Rao Simple Structure (S.S.) Covariance Model.

When S.S. holds $\Sigma = X\Gamma X' + Z\theta Z'$, then the approximate mean of $V^{(2)}$ given $V^{(1)}$ is

$$(2.5) \qquad \mu_{S2\cdot 1} = X^{(2)}T_1F - \Sigma_{S22}^{-1}\Sigma_{S21}(V^{(1)} - X^{(1)}T_1F)$$

where

$$(2.6) \qquad \begin{cases} \Sigma_S = \dfrac{N-r-m-1}{N+3-p+m} MX(X'SX)^{-1}X' + Z(Z'YY'Z)^{-1}Z' = \begin{pmatrix} \Sigma_{S11} & \Sigma_{S12} \\ \Sigma_{S21} & \Sigma_{S22} \end{pmatrix} \\[2ex] T_1 = (X'X)^{-1}X'YA'(AA'). \end{cases}$$

We note that $\Sigma_S$ is obtained from $J$ given by (6.3) of Lee and Geisser (1972) with $t$ being the mode of $F(\cdot; N+1-r-m, N+1-p+m)$, an F distribution.

D) Mode for S.S. Covariance.

The mode of the conditional distribution of $V^{(2)}$ given $V^{(1)}$ when $\Sigma = X\Gamma X' + Z\theta Z'$ satisfies

$$(2.7) \qquad V_m^{(2)} = (E_{22} + \lambda C_{22})^{-1}(E_{22}\mu_{\cdot 2} + \lambda C_{22}\mu_{\cdot 1})$$

where

$$(2.8) \qquad \begin{cases} E = MX(X'SX)^{-1}X' = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} \\[2ex] C = Z(Z'YY'Z)^{-1}Z' = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \\[2ex] \mu_{\cdot 1} = X^{(2)}T_1F - C_{22}^{-}C_{21}(V^{(1)} - X^{(1)}T_1F) \\[1ex] \mu_{\cdot 2} = X^{(2)}T_1F - E_{22}^{-}E_{21}(V^{(1)} - X^{(1)}T_1F_1) \end{cases}$$

From (2.7) we can search for a $\lambda^*$ such that $V_m^{(2)}(\lambda^*)$ maximizes $f(V^{(2)}|V^{(1)})$ where

$$(2.9)\qquad f(V^{(2)}|V^{(1)}) \propto [b_1 + (V^{(2)} - \mu_{.1})'C_{22}(V^{(2)} - \mu_{.1})]^{-(N+1)/2} \cdot [b_2 + (V^{(2)} - \mu_{.2})'E_{22}(V^{(2)} - \mu_{.2})]^{-(N+1-r)/2}$$

and

$$(2.10)\qquad \begin{cases} b_1 = 1 + (V^{(1)} - X^{(1)}T_1F)'C_{11\cdot2}(V^{(1)} - X^{(1)}T_1F) \\ C_{11\cdot2} = C_{11} - C_{12}C_{22}^{-}C_{21} \\ b_2 = 1 + (V^{(1)} - X^{(1)}T_1F)'E_{11\cdot2}(V^{(1)} - X^{(1)}T_1F) \\ E_{11\cdot2} = E_{11} - E_{12}E_{22}^{-}E_{21}\ . \end{cases}$$

The predictive density (2.9) is derived by the authors (1972).

E) Quasi-Least-Squares Predictor (Q.L.S.).

This predictor is defined as $X^{(2)}\hat{\tau}_q$ where $\hat{\tau}_q$ minimizes

$$\left[\begin{pmatrix} V^{(1)} \\ X^{(2)}\hat{\tau}F \end{pmatrix} - X\tau_q\right]' \Omega_q^{-1}\left[\begin{pmatrix} V^{(1)} \\ X^{(2)}\hat{\tau}F \end{pmatrix} - X\tau_q\right]$$

where $\hat{\tau}$ is the MLE of $\tau$,

$$(2.11)\qquad \Omega_q = \begin{pmatrix} \hat{\Sigma}_{11} & 0 \\ 0 & \bar{\Omega}_{22} \end{pmatrix},$$

$\hat{\Sigma}_{11}$ is an estimate of $\Sigma_{11}$ obtained from the sample for

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

which is the covariance matrix of the model considered, and $\bar{\Omega}_{22}$ is the lower right-hand corner of the covariance matrix of the predictive distribution. Hence the Q.L.S. predictor is

$$(2.12)\qquad X^{(2)}\hat{\tau}_q = X^{(2)}(X'\Omega_q^{-1}X)^{-1}X'\Omega_q^{-1}\begin{pmatrix} V^{(1)} \\ X^{(2)}\hat{\tau}_F \end{pmatrix}.$$

We note that in this study we adopt unbiased estimates of $\Sigma$ for $\hat{\Sigma}$ out of the totality of estimators previously given (1972).

F) Simple Least Squares Predictor (S.L.S.).

For this predictor we simply fit a regression based on $V^{(1)}$ without covariance adjustment, i.e., S.L.S. is, when it exists,

$$(2.13)\qquad X^{(2)}\hat{\tau}_s = X^{(2)}(X^{(1)'}X^{(1)})^{-1}X^{(1)'}V^{(1)}.$$

It is clear that for this predictor we take into account $V^{(1)}$ only.

G) $p_1$-Point Predictor ($p_1$- PT.).

For this predictor we estimate $\tau$ if possible by using the first $p_1$ columns only, i.e., $p_1$- PT. is

$$(2.14)\qquad X^{(2)}\hat{\tau}_{p_1} = X^{(2)}(X^{(1)'}\hat{\Sigma}_{11}^{-1}X^{(1)})^{-1}X^{(1)'}\hat{\Sigma}_{11}^{-1}V^{(1)}$$

when it exists.

H) Marginal Predictor (M.P.).

This predictor is obtained by averaging over the simple least squares fits for all individuals and is

$$(2.15)\qquad m_1 = (N-1)^{-1}Le$$

where

$$(2.16)\qquad L = X^{(2)}(X'X)^{-1}X'Y,$$

and $e$ is an $(N-1) \times 1$ vector consisting of all 1's.

I) Ad hoc Predictor (Ad hoc).

This predictor is essentially a combination of two predictors, $m_1$ as defined above and $m_2 = X^{(2)}\hat{\tau}_s$ as given in (2.13). Then the Ad hoc predictor, when it exists, is

$$(2.17)\qquad P_a = (S_{m_1}^{-1} + S_{m_2}^{-1})^{-1}(S_{m_1}^{-1}m_1 + S_{m_2}^{-1}m_2)$$

where $S_{m_1}$ is $\bar{\Omega}_{22}$, as mentioned in (E), when the covariance matrix of the model considered is of Rao's simple structure, and $S_{m_2}$ is the predictive covariance matrix corresponding to $m_2$, i.e.,

$$(2.18)\qquad S_{m_2} = \frac{1}{p_1}[V^{(1)} - X^{(1)}(X^{(1)\prime}X^{(1)})^{-1}X^{(1)\prime}V^{(1)}]'$$
$$\cdot[V^{(1)} - X^{(1)}(X^{(1)\prime}X^{(1)})^{-1}X^{(1)\prime}V^{(1)}][I_{p_2} + X^{(2)}(X^{(1)\prime}X^{(1)})^{-1}X^{(2)\prime}].$$

3. Potthoff-Roy Data.

In order to have some idea of the plausibility of the S.S. model for this set of data we apply the testing statistic as given by Lee and Geisser (1972) to the entire observation set. The design matrices $X$ and $A$ are the same as those used in Khatri (1966), i.e.,

$$X = \begin{bmatrix} 1 & -3 \\ 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix}$$

and $A: 2 \times 27$ composed of 11 $(1, 0)$ columns followed by 16 $(0, 1)$ columns, since the first 11 are girls and the last 16 are boys. In other words we have two groups, each of whose means are linear in time.

The testing statistic is then

$$(3.1)\qquad \lambda_1 = \frac{|(X'S^{-1}X)^{-1}|}{|BSB'|} = 0.8633$$

which is distributed as $U_{2,2,24}$. Hence we have

$$(3.2)\qquad t_1 = \frac{23}{2}\,\frac{1-\sqrt{\lambda_1}}{\sqrt{\lambda_1}} \sim F_{4,46}\,.$$

So $t_1 = 0.88$, which indicates that the null hypothesis $\Sigma = X\Gamma X' + Z\theta Z'$ is not rejected. Thus we shall not rule out the possibility that the covariance matrix is of this particular structure.

We next compare the following predictors:

| Covariance Model | Predictor | |
|---|---|---|
| Arbitrary | Approximate Mean | (A.Me.) |
| | Mode | (Mode) |
| | Quasi-Least-Squares Predictor | (Q.L.S.) |
| | $p_1$ Point Predictor | ($p_1$-PT.) |
| Rao's Simple Structure | Approximate Mean | (A.Me.) |
| | Mode | (Mode) |
| | Quasi-Least-Squares Predictor | (Q.L.S.) |
| | $p_1$ Point Predictor | ($p_1$-PT.) |
| | Marginal Predictor | (M.P.) |
| | Simple Least Squares Predictor | (S.L.S.) |
| | Ad hoc Predictor | (Ad hoc) |

The data of Potthoff-Roy is reproduced in Table 1. In order to compare the predictors we have set $p_1 = 3$ and $p_2 = 1$, i.e., we are thus predicting the last observation. For prediction purposes we withhold one vector and use the rest for predicting the last component of that vector and repeat this for each of the 27 vectors. This gives us 27 predicted values for the 27 last observations. The observations in the last column of Table 1 and the values obtained using the various predictors are presented in Table 2 together with the mean squared deviation (MSD) and the mean absolute deviation (MAD) from the actual observed values. With respect to the MSD, the ad hoc predictor appears to be "best," followed by the mode of $f(V^{(2)}|V^{(1)})$, and then the quasi-least-squares predictor, all based on S.S. model. If individual 20 (a seemingly aberrant vectorial observation) is deleted, then the mode of $f(V^{(2)}|V^{(1)})$ based on S.S. model is best, followed by the ad hoc predictor and the approximate mean of

## TABLE 1

### DENTAL MEASUREMENT OF 11 GIRLS AND 16 BOYS

| *IND/AGE | 8 | 10 | 12 | 14 |
|---|---|---|---|---|
| 1 | 21 | 20 | 21.5 | 23 |
| 2 | 21 | 21.5 | 24 | 25.5 |
| 3 | 20.5 | 24 | 24.5 | 26 |
| 4 | 23.5 | 24.5 | 25 | 26.5 |
| 5 | 21.5 | 23 | 22.5 | 23.5 |
| 6 | 20 | 21 | 21 | 22.5 |
| 7 | 21.5 | 22.5 | 23 | 25 |
| 8 | 23 | 23 | 23.5 | 24 |
| 9 | 20 | 21 | 22 | 21.5 |
| 10 | 16.5 | 19 | 19 | 19.5 |
| 11 | 24.5 | 25 | 28 | 28 |
| 12 | 26 | 25 | 29 | 31 |
| 13 | 21.5 | 22.5 | 23 | 26.5 |
| 14 | 23 | 22.5 | 24 | 27.5 |
| 15 | 25.5 | 27.5 | 26.5 | 27 |
| 16 | 20 | 23.5 | 22.5 | 26 |
| 17 | 24.5 | 25.5 | 27 | 28.5 |
| 18 | 22 | 22 | 24.5 | 26.5 |
| 19 | 24 | 21.5 | 24.5 | 25.5 |
| 20 | 23 | 20.5 | 31 | 26 |
| 21 | 27.5 | 28 | 31 | 31.5 |
| 22 | 23 | 23 | 23.5 | 25 |
| 23 | 21.5 | 23.5 | 24 | 28 |
| 24 | 17 | 24.5 | 26 | 29.5 |
| 25 | 22.5 | 25.5 | 25.5 | 26 |
| 26 | 23 | 24.5 | 26 | 30 |
| 27 | 22 | 21.5 | 23.5 | 25 |

$$S = \begin{pmatrix} 135.39 & 67.92 & 97.76 & 67.76 \\ 67.92 & 104.82 & 73.18 & 82.93 \\ 97.76 & 73.18 & 161.39 & 103.27 \\ 67.76 & 82.93 & 103.27 & 124.64 \end{pmatrix}$$

### DISPLAY OF ESTIMATED VARIANCES, COVARIANCES AND CORRELATIONS

$$\begin{array}{cccc} 5.42 & 2.72 & 3.91 & 2.71 \\ 0.57 & 4.18 & 2.92 & 3.32 \\ 0.66 & 0.56 & 6.46 & 4.13 \\ 0.52 & 0.73 & 0.73 & 4.99 \end{array}$$

*Individuals 1-11 are boys, 12-27 girls.

## TABLE 2

### ACTUAL OBSERVED VALUES AND PREDICTION

| O | ARBITRARY COVARIANCE | | | | RAO'S SIMPLE STRUCTURE COVARIANCE | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | A.ME. | Mode | Q.L.S. | $p_1$-PT. | A.ME. | Mode | Q.L.S. | $p_1$-PT. | M.P. | S.L.S. | Ad Hoc |
| 23 | 22.08 | 22.13 | 22.65 | 20.81 | 22.29 | 22.23 | 22.76 | 21.07 | 24.25 | 21.33 | 21.74 |
| 25.5 | 24.02 | 24.02 | 24.34 | 24.81 | 24.12 | 24.19 | 24.43 | 24.97 | 23.96 | 25.17 | 25.05 |
| 26 | 25.72 | 25.66 | 25.61 | 27.64 | 25.52 | 25.68 | 25.47 | 27.32 | 23.87 | 27.00 | 26.40 |
| 26.5 | 25.89 | 25.88 | 24.77 | 25.92 | 25.93 | 25.90 | 24.80 | 25.88 | 23.87 | 25.83 | 25.82 |
| 23.5 | 24.27 | 24.26 | 23.94 | 23.69 | 24.19 | 24.15 | 23.86 | 23.53 | 24.15 | 23.33 | 23.41 |
| 22.5 | 22.61 | 22.61 | 23.17 | 21.84 | 22.60 | 22.52 | 23.11 | 21.76 | 24.27 | 21.67 | 22.21 |
| 25 | 24.12 | 24.12 | 23.97 | 23.92 | 24.13 | 24.11 | 23.96 | 23.88 | 24.03 | 23.83 | 23.83 |
| 24 | 24.57 | 24.59 | 23.87 | 23.57 | 24.66 | 24.52 | 23.88 | 23.62 | 24.11 | 23.67 | 23.67 |
| 21.5 | 23.17 | 23.19 | 23.76 | 23.00 | 23.14 | 23.10 | 23.70 | 23.00 | 24.30 | 23.00 | 23.00 |
| 19.5 | 21.41 | 21.38 | 23.12 | 21.06 | 21.02 | 21.03 | 22.87 | 20.93 | 24.51 | 20.67 | 21.27 |
| 28 | 27.36 | 27.37 | 26.05 | 28.98 | 27.38 | 27.64 | 26.14 | 29.10 | 23.66 | 29.33 | 28.44 |
| 31 | 28.76 | 28.81 | 27.95 | 28.89 | 28.95 | 28.95 | 28.01 | 29.08 | 27.10 | 29.67 | 28.74 |
| 26.5 | 25.30 | 25.31 | 25.86 | 23.92 | 25.22 | 25.10 | 25.79 | 23.88 | 27.43 | 23.83 | 23.85 |
| 27.5 | 25.53 | 25.56 | 25.76 | 23.78 | 25.53 | 25.36 | 25.78 | 23.94 | 27.38 | 24.17 | 24.42 |
| 27 | 30.03 | 30.00 | 27.74 | 28.09 | 29.40 | 29.23 | 27.55 | 27.81 | 27.33 | 27.50 | 27.47 |
| 26 | 25.85 | 25.80 | 26.50 | 25.19 | 25.50 | 25.48 | 26.28 | 24.90 | 27.44 | 24.50 | 24.62 |
| 28.5 | 28.54 | 28.57 | 27.67 | 28.08 | 28.46 | 28.43 | 27.65 | 28.12 | 27.25 | 28.17 | 28.16 |
| 26.5 | 25.70 | 25.75 | 26.30 | 24.86 | 25.72 | 25.70 | 26.33 | 25.07 | 27.40 | 25.33 | 25.57 |
| 25.5 | 25.26 | 25.35 | 25.22 | 22.51 | 25.39 | 25.25 | 25.49 | 23.15 | 27.48 | 23.80 | 25.26 |
| *26 | 32.74 | 32.26 | 30.62 | 33.67 | 32.18 | 31.22 | 30.84 | 34.66 | 27.27 | 32.83 | 28.47 |
| 31.5 | 31.45 | 31.43 | 29.26 | 31.94 | 31.28 | 31.60 | 29.30 | 32.10 | 27.03 | 32.33 | 31.64 |
| 25 | 25.84 | 25.89 | 25.80 | 23.58 | 25.84 | 25.65 | 25.73 | 23.62 | 27.50 | 23.67 | 23.69 |
| 28 | 26.29 | 26.29 | 26.66 | 25.76 | 26.14 | 26.10 | 26.55 | 25.64 | 27.33 | 25.50 | 25.58 |
| 29.5 | 26.60 | 26.64 | 30.24 | 33.46 | 25.97 | 26.20 | 29.51 | 32.30 | 27.14 | 31.50 | 29.57 |
| 26 | 28.28 | 28.27 | 27.80 | 28.09 | 27.88 | 27.80 | 27.58 | 27.82 | 27.38 | 27.50 | 27.48 |
| 30 | 27.60 | 27.62 | 27.37 | 27.50 | 27.50 | 27.50 | 27.33 | 27.50 | 27.19 | 27.50 | 27.50 |
| 25 | 25.05 | 25.10 | 25.67 | 23.34 | 25.10 | 24.89 | 25.71 | 23.57 | 27.50 | 23.83 | 24.28 |
| MSD1** | 3.734 | 3.462 | 3.057 | 5.536 | 3.375 | 2.877 | 2.970 | 5.509 | 5.417 | 4.000 | 2.110 |
| MSD2** | 2.128 | 2.088 | 2.354 | 3.488 | 2.035 | 1.940 | 2.182 | 2.835 | 5.563 | 2.360 | 1.957 |
| MAD*** | 1.351 | 1.329 | 1.353 | 1.809 | 1.314 | 1.245 | 1.280 | 1.720 | 1.927 | 1.532 | 1.142 |

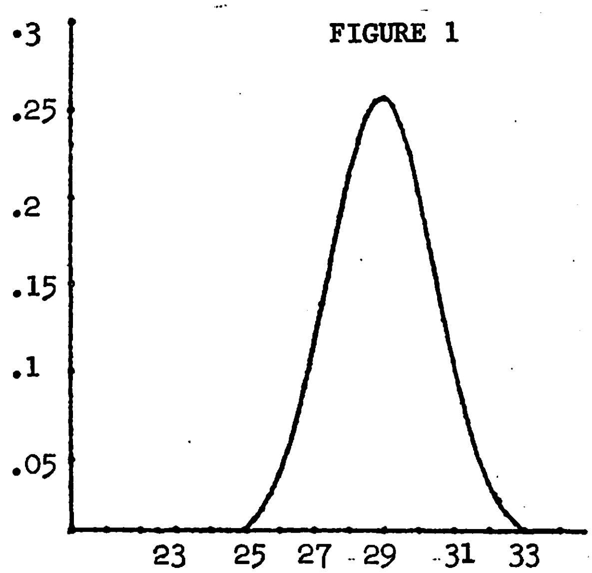
* IND. 20 ** MSD1 is the Mean Sq. Deviation and MSD2 is the Mean Sq. Deviation excluding IND. 20

*** MAD is the Mean Absolute Deviation

$f(V^{(2)}|V^{(1)})$ based on the S.S. model and then the mode and the approximate mean of $f(V^{(2)}|V^{(1)})$ based on the arbitrary covariance. With respect to the MAD, the ad hoc predictor is best, followed by the mode of $f(V^{(2)}|V^{(1)})$, the quasi-least-squares predictor and the approximate mean of $f(V^{(2)}|V^{(1)})$ all based on S.S. model. On the other hand the S.S. modal predictor yields closer values to the actual observations than the ad hoc predictor for about 65% of the individuals.

The exact and approximate p.d.f. of $V^{(2)}$ given $V^{(1)}$ when S.S. model holds are plotted in Figure 1 for $V^{(1)'} = (26, 25, 29)$, i.e., the 12th individual. The approximation given here is the $t$ approximation developed by Lee and Geisser (1972). There is virtually no discernable difference between the exact density and its $t$ approximation.

FIGURE 1




EXACT AND APPROXIMATE $f(V^{(2)}|V^{(1)})$ WHEN S.S. HOLDS

From the above analysis it seems that the mode, the approximate mean of the predictive distribution of $V^{(2)}$ given $V^{(1)}$ and the quasi-least-squares predictor all based on S.S. model as well as the ad hoc predictor, are reasonably good predictors for this data set. Because of the practically perfect fit of the approximate p.d.f. to the exact p.d.f. of $V^{(2)}$ given $V^{(1)}$ as shown in Figure 1, any predictive region obtained from the approximate distribution of $V^{(2)}$ given $V^{(1)}$ based on S.S. model can be expected to yield excellent approximations. In previous analyses of this data set,(Potthof and Roy (1964) and Khatri (1966)), the underlying assumption has been that the covariance matrix was arbitrary. The study here indicates that better results may be attained by assuming that S.S. obtains, at least for prediction purposes.

## 4. Grizzle-Allen Data.

Here, we will again first apply the test for S.S. model. The design matrix $X$ is the same as before while $A$ is a $1 \times 20$ vector consisting of all 1's. Again linearity is assumed.

The testing statistic is

$$\lambda_2 = \frac{|(X'S^{-1}X)^{-1}|}{|BSB'|} = 0.6358 \tag{4.1}$$

which is distributed as $U_{2,2,17}$. Hence we have

$$t_2 = \frac{16}{2} \frac{1-\sqrt{\lambda_2}}{\sqrt{\lambda_2}} \sim F_{4,34} \; . \tag{4.2}$$

Thus $t_2 = 2.03287$, which indicates rejection of the null hypothesis that the covariance matrix is of this particular structure is at approximately the 12 percent level. At this level of significance it would appear that the S.S. model should not necessarily be ruled out as a potential possibility. Hence we will compare the following predictors for several different covariance models:

| Covariance Model | Predictor | |
|---|---|---|
| Arbitrary Covariance | Approximate Mean | (A.ME.) |
| | Mode | (Mode) |
| | Quasi-Least-Squares Predictor | (Q.L.S.) |
| | $p_1$ Point Predictor | ($p_1$-PT.) |

| | | |
|---|---|---|
| Rao's Simple Structure Covariance | Approximate Mean | (A.ME.) |
| | Mode | (Mode) |
| | Quasi-Least-Squares Predictor | (Q.L.S.) |
| | $p_1$ Point Predictor | ($p_1$-PT.) |
| Serial Structure | Quasi-Least-Squares Predictor | (Q.L.S.) |
| | $p_1$ Point Predictor | ($p_1$-PT.) |
| Rao's Factor Structure | Quasi-Least-Squares Predictor | (Q.L.S.) |
| | $p_1$ Point Predictor | ($p_1$-PT.) |
| | Simple Least Squares Predictor | (S.L.S.) |
| | Ad hoc Predictor | (Ad hoc) |

We note that Serial Structure stands for the case where the covariance matrix is of the form

$$\Sigma = \begin{bmatrix} a_1, a_2, \ldots, a_p \\ a_2 \ddots \ddots \ddots \vdots \\ \vdots \ddots \ddots \ddots a_2 \\ a_p, \ldots, a_2, a_1 \end{bmatrix}$$

and Rao's Factor Structure (F.S.), Rao (1967), stands for the case where $\Sigma = C\Gamma C' + \sigma^2 I$, $\Gamma$ is p.d. but unknown, $\sigma^2$ is an unknown scalar, and $C$ is a known matrix. For the F.S. covariance we follow Rao (1967) and Grizzle and Allen (1969), namely, $C$ is assumed to be an arbitrary known matrix. We consider the transformation

$$Y^* = \begin{pmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \end{pmatrix} = \begin{pmatrix} B \\ Z_1' \\ Z_2' \end{pmatrix} Y \tag{4.3}$$

where $Z = (Z_1, Z_2)$. The estimate of the covariance matrix of $Y^*$ is then

$$\hat{\Sigma}^* = \begin{pmatrix} BSB' & BSZ_1 & 0 \\ Z_1'SB' & Z_1'SZ_1 & 0 \\ 0 & 0 & Z_2'SZ_2 \end{pmatrix}. \tag{4.4}$$

We also note that the following four conditions are satisfied in the choice of transformation (1) $B'Z_1 = 0$, (2) $Z_2'X = 0$, (3) $Z_2'C = 0$ and (4) $Z_1'Z_2 = 0$. We thus see that the estimate of the original covariance matrix can be obtained as

$$\hat{\Sigma} = \begin{pmatrix} B' \\ Z_1 \\ Z_2 \end{pmatrix}^{-1} \hat{\Sigma}^*[B, Z_1', Z_2']^{-1}. \tag{4.5}$$

The data of Grizzle-Allen are reproduced in Table 3. In order to compare the predictors we have set $p_1 = 3$, $p_2 = 1$, i.e., we are predicting the last observation. For prediction purposes we withhold one vector and use the rest for predicting the last component of that vector and repeat this for each of the 20 vectors. This gives us 20 predicted values for the 20 last observations. The observations in the last column of Table 3 and the predicted values obtained from the predictors are presented in Table 4A and Table 4B together with the MSD and the MAD from the actual observed values and estimate of $\tau$. With respect to the MSD, the simple least squares predictor is "best," followed by the ad hoc predictor, the approximate mean, the mode of $f(V^{(2)}|V^{(1)})$ and the $p_1$ point predictor all based on S.S. and then the $p_1$ point predictor based on serial covariance and arbitrary covariance. With respect to the MAD, the simple least squares predictor is best, followed by the $p_1$ point predictor based on S.S. and serial covariance and then the mode and the approximate mean of $f(V^{(2)}|V^{(1)})$ based on S.S. model. Comparison among the quasi-least-squares predictors reveals that with respect to the MSD, the predictor based on serial covariance is best, followed by those based on arbitrary covariance, F.S., while with respect to the MAD, the predictor based on arbitrary covariance is best, followed by those based on serial covariance and S.S.

## TABLE 3

### RAMUS HEIGHT OF 20 BOYS

| IND/AGE | 8 | 8.5 | 9 | 9.5 |
|---|---|---|---|---|
| 1 | 47.8 | 48.8 | 49 | 49.7 |
| 2 | 46.4 | 47.3 | 47.7 | 48.4 |
| 3 | 46.3 | 46.8 | 47.8 | 48.5 |
| 4 | 45.1 | 45.3 | 46.1 | 47.2 |
| 5 | 47.6 | 48.5 | 48.9 | 49.3 |
| 6 | 52.5 | 53.2 | 53.3 | 53.7 |
| 7 | 51.2 | 53 | 54.3 | 54.5 |
| 8 | 49.8 | 50 | 50.3 | 52.7 |
| 9 | 48.1 | 50.8 | 52.3 | 54.4 |
| 10 | 45 | 47 | 47.3 | 48.3 |
| 11 | 51.2 | 51.4 | 51.6 | 51.9 |
| 12 | 48.5 | 49.2 | 53 | 55.5 |
| 13 | 52.1 | 52.8 | 53.7 | 55 |
| 14 | 48.2 | 48.9 | 49.3 | 49.8 |
| 15 | 49.6 | 50.4 | 51.2 | 51.8 |
| 16 | 50.7 | 51.7 | 52.7 | 53.3 |
| 17 | 47.2 | 47.7 | 48.4 | 49.5 |
| 18 | 53.3 | 54.6 | 55.1 | 55.3 |
| 19 | 46.2 | 47.5 | 48.1 | 48.4 |
| 20 | 46.3 | 47.6 | 51.3 | 51.8 |

$$S = \begin{pmatrix} 120.27 & 117.59 & 109.76 & 105.42 \\ 117.59 & 122.54 & 116.92 & 112.55 \\ 109.76 & 116.92 & 131.44 & 131.98 \\ 105.42 & 112.55 & 131.98 & 141.83 \end{pmatrix} \tag{4.6}$$

DISPLAY OF ESTIMATED VARIANCES, COVARIANCES AND CORRELATIONS

| | | | |
|---|---|---|---|
| 6.33 | 6.19 | 5.78 | 5.55 |
| 0.97 | 6.45 | 6.15 | 5.92 |
| 0.87 | 0.92 | 6.92 | 6.95 |
| 0.81 | 0.85 | 0.97 | 7.46 |

## TABLE 4A

ACTUAL OBSERVED VALUES AND PREDICTIONS

| O | ARBITRARY COVARIANCE | | | | RAO'S SIMPLE STRUCTURE COVARIANCE | | | |
|---|---|---|---|---|---|---|---|---|
| | A.ME. | Mode | Q.L.S. | $p_1$-PT. | A.ME. | Mode | Q.L.S. | $p_1$-PT. |
| 49.7 | 49.77 | 49.76 | 50.67 | 50.02 | 49.87 | 49.78 | 50.52 | 49.67 |
| 48.4 | 48.56 | 48.56 | 49.90 | 48.61 | 48.59 | 48.54 | 49.85 | 48.39 |
| 48.5 | 48.87 | 48.85 | 49.68 | 48.22 | 48.65 | 48.62 | 49.91 | 48.51 |
| 47.2 | 47.11 | 47.10 | 48.45 | 46.16 | 46.85 | 46.77 | 48.87 | 46.54 |
| 49.3 | 49.75 | 49.74 | 50.57 | 49.82 | 49.77 | 49.72 | 50.49 | 49.60 |
| 53.7 | 54.00 | 54.00 | 52.90 | 54.09 | 54.07 | 54.00 | 52.67 | 53.75 |
| 54.5 | 55.34 | 55.37 | 54.18 | 56.19 | 55.56 | 55.50 | 53.83 | 55.90 |
| 52.7 | 50.91 | 50.91 | 50.82 | 50.49 | 50.79 | 50.89 | 50.94 | 50.54 |
| 54.4 | 52.31 | 52.34 | 53.61 | 55.34 | 53.68 | 53.83 | 53.07 | 54.51 |
| 48.3 | 47.94 | 47.97 | 50.21 | 49.13 | 48.51 | 48.52 | 49.98 | 48.63 |
| 51.9 | 52.43 | 52.41 | 51.64 | 51.80 | 52.25 | 52.14 | 51.64 | 51.80 |
| 55.5 | 53.78 | 53.85 | 53.24 | 54.92 | 53.64 | 53.51 | 53.19 | 54.77 |
| 55 | 54.50 | 54.50 | 53.00 | 54.39 | 54.52 | 54.51 | 53.05 | 54.48 |
| 49.8 | 50.14 | 50.13 | 50.67 | 50.02 | 50.11 | 50.06 | 50.63 | 49.88 |
| 51.8 | 52.10 | 52.09 | 51.75 | 52.00 | 52.07 | 52.06 | 51.75 | 52.00 |
| 53.3 | 53.61 | 53.61 | 52.69 | 53.70 | 53.65 | 53.66 | 52.65 | 53.70 |
| 49.5 | 49.33 | 49.32 | 50.02 | 48.88 | 49.19 | 49.15 | 50.15 | 48.98 |
| 55.3 | 55.93 | 55.94 | 54.38 | 56.61 | 56.09 | 56.08 | 53.91 | 56.09 |
| 48.4 | 49.03 | 49.03 | 50.37 | 49.41 | 49.14 | 49.13 | 50.25 | 49.12 |
| 51.8 | 53.71 | 53.73 | 52.22 | 52.70 | 53.44 | 53.66 | 52.86 | 54.01 |
| MSD | .8583 | .8422 | 1.591 | .7988 | .6972 | .7264 | 1.769 | .7307 |
| MAD | .6791 | .6715 | 1.087 | .7204 | .6344 | .6275 | 1.200 | .5659 |

## TABLE 4B

### ACTUAL OBSERVED VALUES AND PREDICTIONS

| O | SERIAL COVARIANCE | | RAO'S FACTOR STRUCT. | | | |
|---|---|---|---|---|---|---|
| | Q.L.S. | $p_1$-PT. | Q.L.S. | $p_1$-PT. | S.L.S. | Ad Hoc |
| 49.7 | 50.34 | 49.42 | 50.72 | 50.11 | 49.73 | 49.75 |
| 48.4 | 49.71 | 48.24 | 49.92 | 48.67 | 48.43 | 48.45 |
| 48.5 | 49.95 | 48.65 | 49.64 | 48.17 | 48.47 | 48.48 |
| 47.2 | 48.87 | 46.67 | 48.50 | 46.16 | 46.50 | 46.54 |
| 49.3 | 50.36 | 49.43 | 50.60 | 49.88 | 49.63 | 49.64 |
| 53.7 | 52.62 | 53.59 | 52.95 | 54.15 | 53.80 | 53.78 |
| 54.5 | 53.84 | 55.80 | 54.13 | 56.20 | 55.93 | 55.91 |
| 52.7 | 50.90 | 50.59 | 50.85 | 50.46 | 50.53 | 50.53 |
| 54.4 | 52.78 | 54.01 | 53.66 | 55.42 | 54.60 | 54.51 |
| 48.3 | 49.38 | 47.74 | 50.32 | 49.36 | 48.73 | 48.89 |
| 51.9 | 51.64 | 51.80 | 51.65 | 51.80 | 51.80 | 51.80 |
| 55.5 | 53.95 | 56.04 | 52.98 | 54.23 | 54.73 | 54.17 |
| 55.0 | 53.11 | 54.56 | 53.04 | 54.36 | 54.47 | 54.46 |
| 49.8 | 50.54 | 49.78 | 50.69 | 50.05 | 49.90 | 49.90 |
| 51.8 | 51.75 | 52.00 | 51.76 | 52.00 | 52.00 | 52.00 |
| 53.3 | 52.68 | 53.70 | 52.71 | 53.70 | 53.70 | 53.70 |
| 49.5 | 50.13 | 49.04 | 50.01 | 48.86 | 48.97 | 48.97 |
| 55.3 | 53.91 | 55.92 | 54.34 | 56.61 | 56.13 | 56.07 |
| 48.4 | 50.04 | 48.83 | 50.42 | 49.52 | 49.17 | 49.19 |
| 51.8 | 53.10 | 54.28 | 51.87 | 52.10 | 53.40 | 53.21 |
| MSD | 1.523 | .7356 | 1.672 | .8900 | .6405 | .6691 |
| MAD | 1.122 | .5711 | 1.091 | .7665 | .5650 | .5829 |

From the above analyses it seems that the simple least squares predictor, the ad hoc predictor and the $p_1$ point predictor are good predictors for this data set.

The above study does not indicate clearly which covariance model is most appropriate for this data set for the purposes of prediction, since the "best" predictors are not necessarily related to a particular model. However, the serial model is suspected of being most appropriate here but unfortunately predictive density results are not yet available for this model.

## 5. Concluding Remarks.

From this empirical study, there is some support for the robustness of the ad hoc predictor in the sense that no matter what the perturbation from the real model (provided it is not described), this procedure seems to provide reasonable predictions. In case the real model is known, then better predictors will undoubtedly be those obtained under that particular model. We thus tentatively recommend the following procedure in predicting growth data: First of all, apply the testing statistic to see if the S.S. is tenable. If the null hypothesis is not rejected, then predictors corresponding to this model as well as the ad hoc predictor should be computed. If the ad hoc predictor turns out to be best either use it or conduct a search for a more appropriate covariance pattern. On the other hand, if the ad hoc predictor is considerably poorer than those based on S.S. model, one ought to choose the best among the predictors based on that structure. If the null hypothesis does not appear to be tenable, then the predictors corresponding to the arbitrary covariance as well as the ad hoc predictor should be computed and again if the ad hoc predictor is best a search for a suitable pattern is in order if possible, and the predictors should be modified to take into account whatever pattern seems appropriate. If the ad hoc predictor is much worse and no pattern apparent, then choose the best amongst the predictors based on the arbitrary covariance. A predictive region can also be obtained by using a normal or $t$ approximation depending on the covariance structures.

The ad hoc predictor, it is to be noted, has the drawback that it will not always exist. On the other hand, under most circumstances, the quasi-least-square predictor is rather easy to compute, as compared with the first four conditional predictors and may lead to reasonable predictions. Besides, this method can also be used to extend predictions to the $p+1^{th},\ldots, p+\ell^{th}$ components of a given individual when the correlation pattern between $p+1^{th},\ldots, p+\ell^{th}$ and the whole $p$ vector has an ascertainable pattern. For example, if $\Sigma_{4\times 4}$ is of the pattern

$$\Sigma = \sigma^2\begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix},$$

then from the sample covariance matrix we can estimate $\sigma^2$ and $\rho$ as $\hat{\sigma}^2$ and $\hat{\rho}$. If we are interested in predicting the $5^{th}$ time point for any individual, say $Y_1$ for $Y_{4\times N} = (Y_{.},\ldots, Y_N)$, then we can use the quasi-least-square predictor

$$\tilde{X}\hat{\tau}_q = \tilde{X}(X^{*'}\Omega_q^{-1}X^{*})^{-1}X^{*'}\Omega_q^{-1}\binom{Y_1}{\tilde{X}\hat{\tau}A_1}$$

where

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix},$$

$$\tilde{X} = (1,\ 5),$$

$$X^{*} = \binom{X}{\tilde{X}},$$

$$\Omega_q = \hat{\sigma}^2 \begin{pmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & \hat{\rho}^3 & \hat{\rho}^4 \\ \hat{\rho} & 1 & \hat{\rho} & \hat{\rho}^2 & \hat{\rho}^3 \\ \hat{\rho}^2 & \hat{\rho} & 1 & \hat{\rho} & \hat{\rho}^2 \\ \hat{\rho}^3 & \hat{\rho}^2 & \hat{\rho} & 1 & \hat{\rho} \\ \hat{\rho}^4 & \hat{\rho}^3 & \hat{\rho}^2 & \hat{\rho} & 1 \end{pmatrix},$$

and

$$A = (A_1, \ldots, A_N).$$

The central purpose of this paper was to illustrate the computational feasibility of certain prediction methods on real data sets for a particular prediction problem. In accomplishing this we have also tacitly suggested a data analytic technique for providing comparisons among competitive predictors. More formally we now propose that if a data analyst is interested in prediction he should avail himself of the opportunity of sorting alternatives on the data set itself.

In our example we assumed linear growth curve and searched for the best predictors under various covariance patterns. Certainly the method could just as well have been turned around to determine whether the growth curves were linear or some other polynomial in time; the logic of course being that the situation which renders the best prediction is mostly likely to obtain. Hence we maintain that this approach can be a valuable guide in determining generally which of a variety of models may be best suited for the data at hand even for purposes other than prediction. The term "models" not only includes the structure of the likelihood itself but also those elusive Bayesian prior assumptions. With regard to the latter, Bayesians are prone to either subjective speculation or mathematical convenience coherent or otherwise, rather than data analytic justification.

The computations reported in this paper were done on APL terminals at State University of New York at Buffalo and the University of Minnesota.

References.

[1] Geisser, S. (1970). Bayesian analysis of growth curves. Sankhyā A 32 53-64.

[2] Grizzle, J. E. and Allen, D. M. (1969). Analysis of growth and dose response curves. Biometrics 25 357-381.

[3] Khatri, C. G. (1966). A note on MANOVA model applied to problems in growth curve. Ann. Inst. Stat. Math. 18 75-86.

[4] Lee, J. C. and Geisser, S. (1972). Growth curve prediction. (To appear in Sankhyā A 34).

[5] Potthoff, R. R. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 51 313-326.

[6] Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proc. Fifth Berkeley Symp. on Math. Stat. and Prob. 1 355-372.